Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Environmental Sciences 7 (2011) 146–151

Procedia Environmental Sciences

Spatial Statistics 2011

Disaggregating gridded air quality data for individual exposure modelling

Lydia E. Gerharz[a*], Benedikt Gräler[a], Edzer Pebesma[a]

[a]*Institute for Geoinformatics, University of Muenster*

Abstract

This paper presents an analysis of disaggregation for PM_{10} pollution from a grid to point support for exposure modelling on a GPS track representing an individual space-time trajectory. Different sets of explanatory variables were tested to predict spatial variability of mobile PM_{10} measurements at the point support. Disaggregation was performed using unconditional Gaussian simulation. The results show a considerable amount of uncertainty added due to disaggregation that depends in strength on the auxiliary data set used for prediction. Subsequent aggregation over the GPS track leads to a reduction in uncertainties.

Keywords:Air pollution, uncertainties, health effects, GPS tracks, Change of support, Unconditional simulation

Nomenclature

PM_{10}	Airborne particulate matter with aerodynamic diameter smaller than 10 µm

1. Introduction

Exposure modelling at the individual level requires the integration of aggregated air quality information such as model results on a grid with a trajectory of individual locations obtained from GPS

* Corresponding author. Tel.: +49 251 83 39760; fax: +49 251 83 39763.
E-mail address: Lydia.gerharz@uni-muenster.de

doi:10.1016/j.proenv.2011.07.026

records. This Change of Support problem is well known in environmental and exposure science [1] but has to our knowledge not been addressed on the level of individual space-time tracks so far. Typically, air quality data needs to be downscaled or disaggregated from the larger spatial units to overcome the mismatch of spatial and temporal support. This introduces additional uncertainty for estimates on point support. Disaggregation can be done by unconditional simulation when no point measurements are available and requires knowledge of the spatio-temporal variability at point support.

The goals of the presented study are to develop a model for spatial variability by relating variogram coefficients to external, readily available information such as weather and land use. This should ensure transferability ofapproach to other areas by using such information for prediction of spatial variability.

Based on a set of mobile measurements of particulate matter smaller than 10 µm in diameter we will develop a model to predict point support variability of air pollution using auxiliary data. This model can be used for disaggregation of grid support PM_{10} data where no point measurements are available. A number of different spatial and temporal predictors will be presented and compared with respect to their influence on the uncertainty of an exposure trajectory. Influence of subsequent aggregation of the point estimates on the uncertainties will be demonstrated to show the relevance of addressing disaggregation methods in exposure modelling studies.

2. Data

The present study examines 16 tracks of PM_{10} measurements (covering approximately 2 hours each) that were collected between March 22 and April 13, 2010 in the city of Münster, Germany. PM_{10} mass concentrations in $\mu g\ m^{-3}$ as well as the position in WGS84 coordinates were measured each second using a TSI DustTrak DRX Handheld and a Holux GPS tracker M-241.

Auxiliary data sets for further analysis of the mobile measurements are shown in table 1. Whereas CLC, StreetDensity and TrafficEmissionDensity can be considered as spatially variable on the grid scale and temporally static, Meteorology and TrafficCount are spatially static and temporally variable. AustalPM10 are spatially and temporally variable results from the Lagrangian air pollution dispersion model AUSTAL2000 [2].

The disaggregation simulation uses a GPS track collected by a test person. Each track covers an outdoor travel of approximately 1 hour in the city of Münster with positions recorded every second.

Table 1. Auxiliary data sets used for PM_{10} track analysis.

Data set	Description	Spatial resolution	Temporal resolution
CLC	CORINE land cover data	250 x 250 m²	Static for 2000
StreetDensity	Line density calculated from street polylines	250 x 250 m²	Static, actuality varies
TrafficEmissionDensity	Line density for annual traffic PM_{10} emissions from major streets	250 x 250 m²	Annual for 2005
AustalPM10	PM_{10} concentration modelled with AUSTAL2000	250 x 250 m²	1 hour
Meteorology	Wind speed and direction from a measurement station in Münster	1 observation for Münster	10 min
TrafficCount	Average traffic count for working days estimated from 46 sampling days at 13 different locations	Average over Münster	1 hour, annual average

3. Methodology

For disaggregation of spatio-temporal data from a grid, additional information to infer the point support variability σ^2 of the phenomenon is necessary [3]. The grid used is the AUSTAL2000 air pollution prediction grid with 250x250 m² resolution (see table 1). As air pollution is a field phenomenon and thus spatially and temporally correlated, the strength of autocorrelation $\rho(h)$ needs to be taken into account as a second factor for disaggregation.

The (semi-)variogram is used to represent the spatial variability of PM_{10} at point support. The variogram $\gamma(h)$ describes the semi-variance of the values for the variable Z between two locations separated by the distance h. It is closely related to the variance σ^2 and the autocorrelation $\rho(h)$ of Z, as $\gamma(h)$ for large distances equals σ^2 and the strength of the autocorrelation determines how the semi-variance increases with distance h:

$$\gamma(h) = \sigma^2 - \sigma^2 \rho(h) \tag{1}$$

For the estimation of the variogram, a number of PM_{10} measurement tracks were used. Based on explanatory analysis we decided to choose an exponential variogram model with nugget C_0, partial sill C_1 and range a, for our study:

$$\gamma(h) = \begin{cases} 0 & if\ h = 0 \\ C_0 + C_1\left(1 - \exp\left(\frac{-h}{a}\right)\right) & if\ h > 1 \end{cases} \tag{2}$$

Exponential variogram models were fitted automatically (using the automap package in R) for each 250x250 m² grid cell with more than 60 measurement points. Outliers were removed manually during the estimation of exponential variogram models per grid cell.

To estimate the variogram parameters at un-sampled locations, a set of auxiliary variables (see table 1) were tested for correlation and linear regression models were fitted to σ, C_0, C_1 and a per grid cell. For estimation of the variogram parameters C_0, C_1 and a, multiple linear regression was applied, using data sets that showed significant fits of the linear regression model. The disaggregation itself was obtained by unconditional Gaussian simulation. The estimated variogram models of each grid cell were used for unconditional Gaussian simulation for points of a sample GPS track. The results were averaged over the number of points visited per grid cell in order to compare degree of uncertainty reduction due to aggregation. Overall track results were compared between different regression models to test robustness of the used method.

All analyses were performed using the statistical software R [4] and the R packages gstat [5] and automap [6].

4. Results

In table 2 the rank correlation coefficients for valid variogram parameters with the auxiliary data sets are shown for each grid cell. Strong correlations can be found between variogram parameters, with exception of the range, and wind direction and between all variogram parameters and CORINE land cover data. The range parameter shows the weakest correlation with the auxiliary data sets for the variogram parameters.

Table 2. Spearman's rank correlation coefficients for auxiliary data sets and variogram parameter and standard deviation of PM_{10} point measurements per grid cell.

Auxiliary data set	nugget	Partial sill	range	Standard deviation
Meteorology: Wind direction	-0.22	-0.38	-0.04	-0.37
Meteorology: Wind speed	0.15	0.26	0.17	0.18
TrafficCount	-0.07	0.02	-0.18	0.11
AustalPM10	-0.10	-0.09	-0.20	-0.04
CLC	0.23	0.20	0.17	0.10
TrafficEmissionDensity	0.16	0.23	0.07	0.21
StreetDensity	0.14	0.14	0.04	0.11

For all parameter and data set combinations shown in table 3, linear regression models were fitted. Models with significant coefficients were used to build the prediction model SIG for the variogram parameters shown in table 3. In the SIG prediction both, temporal and spatial, data sets were used for parameter estimation. For further comparison, four other prediction methods were included in the analysis as shown in table 3. SPAT uses spatially variable and TEMP temporally variable predictor data sets only. ALL includes all auxiliary variables from table 1, regardless of the correlation strengths. A general approach is shown in the GEN prediction method where the partial sill is approximated by the within-cell standard deviation, range is assumed to be constant and averaged over the area and the nugget is considered to be zero. These rough estimates form a model that is easily transferable to other regions without further data requirements.

Results for unconditional simulation on the sample GPS track using the SIG prediction parameters with 10,000 simulations were averaged over the points per grid cell. Box plots of the simulation averages are given in figure 1. Averaging over more than one point per grid cell already reduces the uncertainty between simulations. However, this effect seems to weaken for more than 50 points per cell and is, furthermore, also dependent on the variogram parameters, which vary per grid cell. This leads for example to a larger interquartile range for the last cell when averaging takes place over 24 points compared to the sixth cell with averaging over 12 points only.

The comparison of results from unconditional simulation using the five different prediction models is shown for the whole GPS track in figure 2. ALL shows the largest variability, GEN the smallest. The selected approach, SIG, which only uses significant linear regression predictor variables shows average variability compared to other approaches. It is clear from figure 2, that using different data sets for the prediction model may lead to different estimates of the uncertainty due to disaggregation.

Table 3. Auxiliary data sets used in linear regression models for prediction of variogram parameters

	SIG	ALL	SPAT	TEMP	GEN
Partial sill	wind direction, wind speed, CLC	all	CLC	wind direction	wind speed, StreetDensity
Range	wind speed, AustalPM10	all	AustalPM10	wind speed	Average from all variograms
Nugget	wind direction, CLC	all	CLC	wind direction	

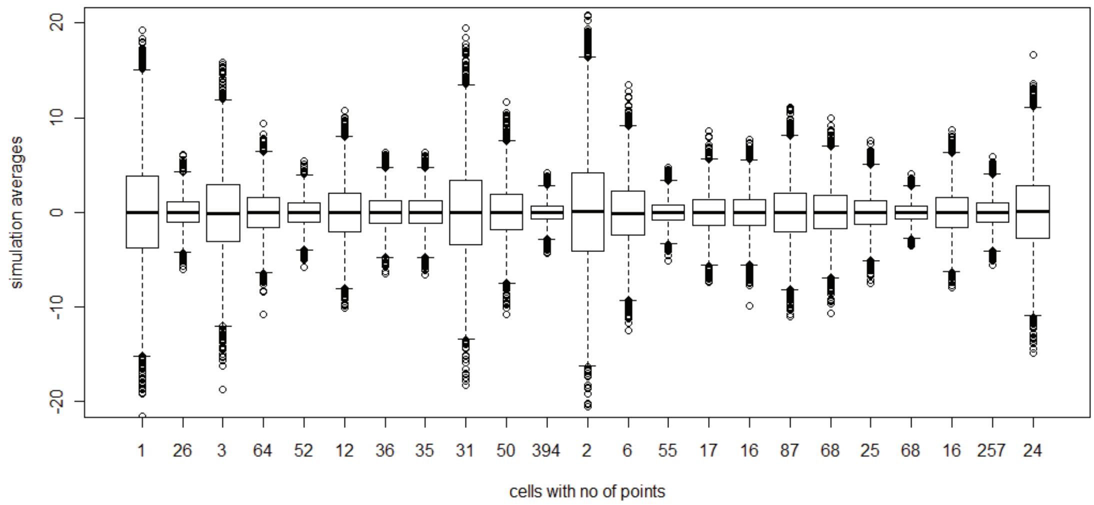

Figure 1. Distribution of disaggregation simulation averages per grid cell for the SIG model with number of points per cell.

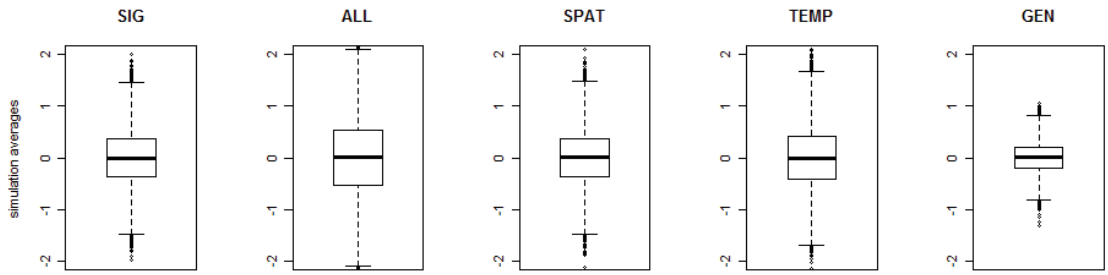

Figure 2. Distribution of disaggregation simulation means over the whole track using different predictor models.

5. Discussion and Conclusion

In this study a methodology for disaggregation of PM_{10} concentration from grid to point support for exposure modelling on a GPS track was developed and applied. We did obtain models that predict spatial variability of PM_{10} at the point support from readily available variables such as weather and land use. An analysis of point support PM_{10} measurements for variogram estimation was shown. Different auxiliary data sets for prediction of the variogram model parameters were tested. Results showed sensitivity to the selection of data sets used for prediction and a substantial amount of uncertainty introduced by this disaggregation. Subsequent aggregation of the points over cells or even over the whole track clearly reduced the uncertainty again. The strength of this effect is determined by the strength of spatial autocorrelation. Therefore, spatial autocorrelation should be included in proper disaggregation and aggregation steps in exposure modelling. We conclude that methods to assess the spatio-temporal autocorrelation and covariates for prediction are essential to reduce uncertainties in individual exposure modelling.

A number of assumptions and limitations were necessary to perform this study. The varying prediction capabilities of the auxiliary data sets led to differences in the disaggregation results as the comparison of the methods showed. This may also be due to unpredictable measurement errors like sensitivity of the measurement device accuracy to air humidity due to the optical measurement technique. The assumption of an exponential variogram and neglecting anisotropy in the study area may be limitations with a rather weak influence compared to the difficulties in predicting the variogram parameters correctly.

Acknowledgements

This work was partially funded by the EU projects HEIMTSA (FP6-036913) and UncertWeb (FP7-248488). The views expressed herein are those of the authors and are not necessarily those of the European Commission.

References

[1] Gotway CA, Young LJ. Combining incompatible spatial data. *J Am Stat Assoc* 2002;**97**:632–48.

[2] Janicke L, Janicke U.: The development of the dispersion model AUSTAL2000G. Reports on Environmental Physics No. 5, 2002, ISSN 1439–8303 (German). Provided at http://www.janicke.de.

[3] Bierkens MFP, Finke PA, de Willigen P. *Upscaling and Downscaling methods for Environmental Research*. Doordrecht: Kluwer Academic Publishers; 2000.

[4] Ihaka R, Gentleman R. R. A Language for Data Analysis and Graphics. *J Comput Graph Stat* 1996;**5**:299–314.

[5]Pebesma EJ. Multivariable geostatistics in S: the gstat package. *Comput Geosci* 2004;**30**:683-91.

[6] Hiemstra PH, Pebesma EJ, Twenhöfel CJW, Heuvelink, GBM. Real-time automatic interpolation of ambient gamma dose rates from the Dutch Radioactivity Monitoring Network. *Comput Geosci* 2009;**35**:1711-21.